

# OCR Indexing

Since Tiki20, file galleries can index the contents of files with images uploaded to Tiki, by means of "Optical Character Recognition" (OCR), and take the result to feed the search index also.

Tiki relies on <https://github.com/tesseract-ocr/tesseract> so you need to install as per <https://tesseract-ocr.github.io/tessdoc/Installation.html>

If you are using WikiSuite, Tesseract is installed by default:

<https://wikisuite.org/Differences-between-Virtualmin-and-WikiSuite>

Server Check helps you confirm that Tesseract is working well, and available to Tiki.

# Required Preferences

To enable OCR indexing in Tiki, make sure to activate the following preference:

*ocr\_enable*: Enables Tiki to extract and index text from supported file types.

# Optional Preferences

You can further customize OCR behavior with these optional settings:

*ocr\_every\_file*: If enabled, Tiki will attempt OCR on all supported files, regardless of other criteria.

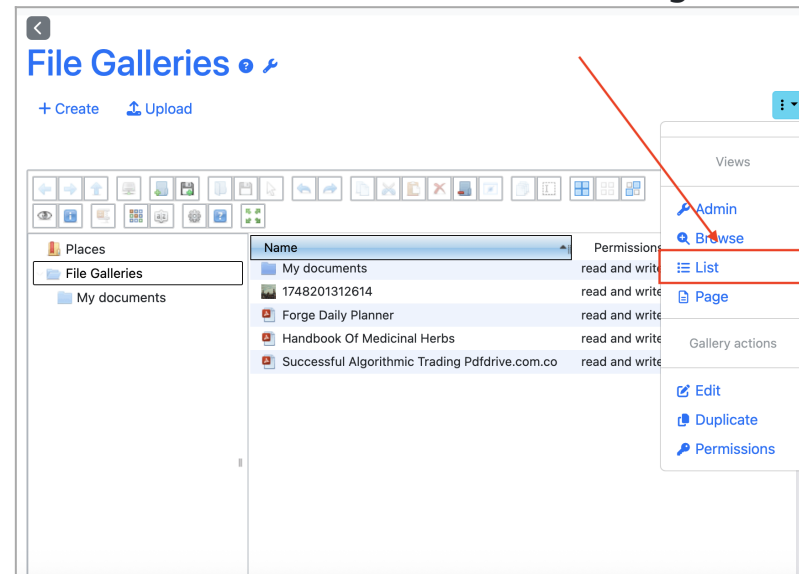
*ocr\_file\_level*: Allows users to override the default OCR language settings on a per-file basis.

# Additional Customization

Tiki also offers several advanced customization options:

- Display OCR status per file.
- Set custom paths for the tesseract and pdftimages binaries via the system \$PATH.

The file gallery has two view modes: "Finder view" and the default "List view". The OCR status for files can only be seen in the default "List view" mode of the file galleries File Gallery.

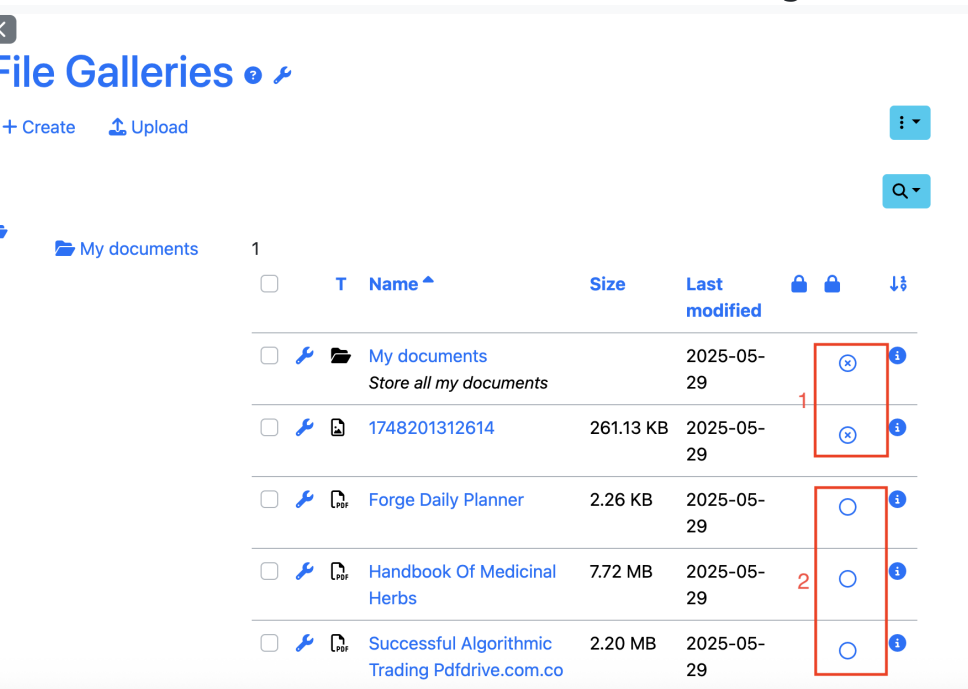


This is where you can switch between File Galleries modes — from the default list view to Finder view, and vice versa.

# Understanding OCR Status in File Galleries

When OCR is enabled, Tiki displays the processing status of each file directly in the File Galleries list. This helps users know whether a file's text content has been indexed and is searchable.

The image below shows few OCR statuses in action:



## (1) No scheduled processing:

This status means the file hasn't been marked for OCR yet. This could be because the file type isn't supported, or the file hasn't met criteria for OCR processing.

## (2) Queued for processing:

The file has been added to the OCR queue and will be processed shortly. Once done, the status will change accordingly.

Other possible statuses include:

### Finished processing

→ OCR has successfully extracted text and indexed the file. The file is now searchable by its content.

### Currently processing

→ Tiki is actively running OCR on this file.

### Processing stalled

→ An error occurred or the process is stuck. This may require admin review.

Alias names for this page:

OCR | OCRIndexing | Optical Character Recognition